Journal of Soft Computing and Computer Applications

Volume 2 | Issue 1 Article 1016

2025

Arson Event Detection Using YOLOv9

Ali Abbas Abbod
University of Technology – Iraq, Department of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq, cs.22.19@grad.uotechnology.edu.iq

Matheel E. Abdulmunimb
University of Technology – Iraq, Department of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq, Matheel.E.Abdulmunim@uotechnology.edu.iq

Ismail A. Mageed
Sheffield Institute of Education, Sheffield Hallam University, Sheffield, United Kingdom, c4031056@hallam.shu.ac.uk

Recommended Citation

Abbod, Ali Abbas; Abdulmunimb, Matheel E.; and Mageed, Ismail A. (2025) "Arson Event Detection Using YOLOv9," *Journal of Soft Computing and Computer Applications*: Vol. 2: Iss. 1, Article 1016.
DOI: https://doi.org/10.70403/3008-1084.1016

JOURNAL OF SOFT COMPUTING AND COMPUTER APPLICATIONS 2025;2:1016

ORIGINAL STUDY

Arson Event Detection Using YOLOv9

Ali Abbas Abbod [a,*], **Matheel E. Abdulmunimb** [a], **Ismail A. Mageed** [b]

[a] University of Technology – Iraq, Department of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq
[b] Sheffield Institute of Education, Sheffield Hallam University, Sheffield, United Kingdom

ABSTRACT

Detecting event anomalies is crucial for surveillance systems, as it enables the identification of occurrences in videos, both temporally and spatially. It can identify deviations from patterns without requiring human oversight by learning from past information to distinguish normal behavior and pinpoint irregularities. Early detection of arson fires is critical to mitigating damage, public safety, property, and the environment, as well as saving lives and aiding in law enforcement investigations. The objective of this study is to evaluate a system for detecting events using the You Only Look Once version 9 (YOLOv9) model in surveillance videos with a focus on identifying incidents of arson. This involved gathering and organizing data, adding annotations, enhancing the dataset through model training methods, assessing the model's performance, and comparing results while considering the dataset quality and diversity as well as environmental factors. In the arson category, the model scored 0.552 in precision, which means there is a tradeoff between precision and recall with threshold 0.5. Additionally, it demonstrated a precision of 1.00 with a confidence of 0.933, meaning it can make predictions with certainty. Finally, the results demonstrate that the model is capable of detecting arson in surveillance systems. The next step will be to broaden the scope of data and improve the model to make it more effective and reliable in other conditions and scenarios.

Keywords: Anomaly detection, Deep learning, YOLOv9, Convolutional neural networks

1. Introduction

Anomaly events are deviations from expected behavior, which can be deviations from the expected value of individual data points, contextual irregularities, or strange patterns in the data [1–3]. Intelligent surveillance systems depend on anomaly event detection in order to detect anomalous video events temporally and spatially [4]. The conventional approach requires a certain number of people to be watched and prone to human mistakes.

Received 12 July 2024; revised 28 March 2025; accepted 11 April 2025.
Available online 14 June 2025

* Corresponding author.
E-mail addresses: cs.22.19@grad.uotechnology.edu.iq (A. A. Abbod), matheel.e.abdulmunim@uotechnology.edu.iq (M. E. Abdulmunimb), c4031056@hallam.shu.ac.uk (I. A. Mageed).

https://doi.org/10.70403/3008-1084.1016

Machine learning (ML) and deep learning (DL) algorithms have automated systems that have revolutionized the field substantially, and they have greatly improved anomaly detection accuracy and efficiency while also making systems much more reliable [5]. The intelligence capabilities of automated systems for detecting anomalies include processing large volumes of video data instantaneously [6]. However, these systems can also learn from past information to find deviations, especially abnormalities, without human intervention [7, 8]. This technological progress is of value in demanding settings such as airports, train stations, shopping centers, and vital facilities, where continuous surveillance is essential to ensure well-being. Arson is the use of fire or arson devices to intentionally damage or attempt to damage any personal or social property. Arson is an early detection that is important to reduce damage, public safety, property, and the environment, as well as to save lives and aid law enforcement investigations [9]. Compared to the traditional thermal sensors for fire detection, artificial intelligence models such as You Only Look Once (YOLO) are well known for their high speed and accuracy in object recognition in videos and images [10]. One of the latest versions in this series, You Only Look Once version 9 (YOLOv9), shows notable progress over previous models, such as YOLOv5 and YOLOv8, and it is superior to past models like Faster Region-Based Convolutional Neural Network (Faster R-CNN) and single shot detectors (SSDs) in speed and accuracy [11, 12]. It is able to process video in real time at more than 60 frames per second (FPS) with an accuracy of more than 90%, which makes it very effective at detecting fires in low light conditions or bad weather [13]. Apart from that, YOLOv9 can minimize the false alarm rate by utilizing advanced pattern analysis and object detection techniques [14]. Because of this, YOLOv9 is a suitable technology for arson fire detection in surveillance systems to improve response speed and protect lives and property. Detecting arson fires early is vital for the following reasons [15]:

- Fast arson detection enables emergency responders to reach the scene quickly, thus lowering the risk of injuries and fatalities and decreasing property and infrastructure damage.
- Arson cases are detected and recorded precisely, which provides evidence for the investigation and helps law enforcement agencies to identify and arrest the culprits.
- Prompt action can prevent fires from spreading, thus safeguarding wildlife and ecosystems.

Advanced models like YOLOv9 in arson fire detection systems can greatly increase their capability of firing fire detection at precise locations, even in challenging conditions. YOLOv9 is an object detection model which is known to be fast and accurate in detecting objects in images or video frames. It is promising for application in arson fire detection to create more effective and reliable monitoring systems [16].

The purpose of this study is to build and test a system for spotting activities using the YOLOv9 learning model in surveillance videos that focus on arson incidents. Arson fires are a concern for damage, loss of life and serious economic harm. Setting arson to property as a means of destroying and causing harm to the public safety and environment is a serious threat to public safety and the environment. This study applies the YOLOv9 model to detect arson in video surveillance systems. It includes data collection, model training, performance evaluation and the comparative analysis with the available detection methods. Furthermore, the study covers dataset quality and diversity, environmental conditions, and a future research suggestion to facilitate real time implementation and integration with monitoring systems.

The main contributions of this study are as follows:

1. Create a dataset of arson from an original video dataset to serve as an essential reference for arson research.
2. Use YOLOv9 to improve the accuracy of detecting arson incidents and enhance response times.
3. Modify and train the YOLOv9 model on a wide range of arson-related events.
4. Programmatic methods were used to split and augment the data using a Python script.

This study is organized as follows: Section 2 reviews research about anomaly detection methods and the development of YOLO, along with various fire detection approaches. Section 3 details implementation steps, starting with data labeling and augmentation methods, followed by YOLOv9 model training procedures and the description of the issues encountered and the utilied solutions and model enhancement methods. Section 4 presents the quantitative model performance outcomes, including precision, recall, and fire recognition accuracy measurements. Section 5 describes the study's primary results and contributions while examining public security applications and proposing future research directions.

2. Related works

This section discusses studies on detecting anomalies in video surveillance using learning algorithms. It highlights the advancements made with the YOLO algorithm, leading to the development of YOLOv9 as an object detection model. The related works are organized into two main categories.

2.1. Detecting anomaly event using deep learning algorithms

Virender et al. [17] presented rapid method for determining whether an unexpected action is unusual or suspicious. It is necessary to indicate which frames and segments of the recording feature the unusual activity. Thus, they used DL techniques to automate the threat recognition system to minimize the waste of labor and time. It aims to distinguish abnormalities from typical patterns by recognizing aggressive and violent indicators in real time. They plan to apply two distinct DL models, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), to detect and categories high movement levels in the frame. Subsequently, they were able to detect warnings in the event of a threat, highlighting any suspicious activity at that particular moment.

Abhishek et al. [18] propose an approach to detect weapons in time using CNN methods alongside the YOLOv9 object recognition system in both live and recorded videos. By incorporating YOLOv9, the authors have significantly enhanced the accuracy and speed of weapon detection, enabling the identification of threats. This method demonstrates performance across lighting conditions and environments, showcasing high recall rates and precision through rigorous testing and evaluation. Leveraging CNN-based architecture and learning techniques, they have effectively done this. classified weapons in video frames, with an accuracy rate of 97.62%.

Pushpajit and Praveen [19] presented a framework based on semi-supervised DL to detect anomalous events in sensitive environments such as Automated Teller Machines (ATM). The framework used multimodal data and achieved competitive results with state-of-the-art methods on benchmark datasets such as Avenue and UFCrime2Local. The framework requires only regular video samples for training, which reduces computational

complexity. A new dataset has been collected to enhance accuracy and security at real monitoring sites.

Sardar et al. [20] showed detecting traffic accidents through surveillance videos using DL techniques. Statistically, Violent traffic incidents occurring in real world traffic scenarios have been increasing and therefore, proposed an attempt of an automatic accidents detection methodology based on CNNs given massive use of Video Traffic Surveillance Systems (VTSS). In order to help support the training process, they built a specialized dataset called Vehicle Accident Image Dataset (VAID) consisting of 1360 labelled accident images. Since the dataset size is small, rotation, shear and zoom and flip were applied to augment the data. To solve for prediction instability (label flickering), a rolling prediction average algorithm was applied to the trained CNN model, and 30 traffic surveillance video were used for validation. They were able to achieve 82% in terms of accuracy on high resolution videos for detecting traffic accident events. Although, the approach was less effective in videos with distant views or bad visibility, for example in fog. The achieved results affirm the usefulness of DL in reducing the delay to response time to traffic accidents and helping monitoring personnel through educated video analysis.

Rida et al. [21] presented the critical issue of insider threats, which often surpass external attacks in terms of damage due to the attacker's legitimate access and ability to bypass conventional security measures. This study suggests a DL based behavioral analysis framework to identify symptom signs of insider threat that are anomalous user activities. Based on the user generated event logs such as logon/logoff events, user roles and functional units, the system creates the rich behavioural features. The authors trained their model using the CMU CERT v4.2 insider threat dataset and implemented a Long Short-Term Memory (LSTM) autoencoder to distinguish between normal and malicious behavior. Evaluation results showed that the proposed model outperformed some of the benchmark models, including LSTM-CNN, Random Forest (RF), Markov chain model and one-class Support Vector Machine (SVM), attaining an accuracy of 90.60%, precision of 97.00% and F1-score of 94.00%. Emphasizing the need of behavioral context in anomaly detection, this method shows to be efficient in spotting insider threats with little domain knowledge and computational load.

Chao et al. [22] use ML and DL for improving the accuracy, efficiency of network anomaly detection in a hybrid Intrusion Detection System (IDS). A scalable, two stage framework in which first distributed K-means and distributed RF algorithms are run on the Spark platform for fast binary classification of normal and abnormal events. In the second stage, DL techniques such as CNN and LSTM are applied to further classify the detected anomalies by specific attack types. Adaptive Synthetic Sampling (ADASYN) is applied during training to overcome the problem of class imbalance. With regard to evaluation using the NSL-KDD and CIC-IDS2017 datasets, the high classification performance resulted in an accuracy rate of 85.24 in NSL-KDD and 99.91 in CIC-IDS2017 respectively. Results show that the model is competent to handle broad attack types, faster preprocessing and streamlined training time than DL models.

2.2. Analyzing and classifying arson and fire data using machine learning

Park et al. [23] presented a real-world Gas Chromatography–Mass Spectroscopy (GC–MS) dataset with around 4000 suspected arson cases, and three classification models based on ML were created. These models were trained to categorize fire residue data into six groups. The random forest, support vector machine, and CNN models achieved classification accuracies of 0.88, 0.88, and 0.92, respectively.

Nikolay et al. [24] presented the identification of smoke and fire in a given area using drones. The research aims to enhance the quality of video data by reducing motion blur, stabilizing video streams, detecting the horizon line in frames and recognizing fires through segmentation techniques involving Euclidean–Mahala Nobis distance and a modified YOLO CNN. The horizon line detection algorithm proposed helps eliminate details like cloud-covered regions by assessing contrast, which indicates pixel informativeness. By employing these methods, the system introduces a slight delay of 0.03 seconds due to its efficient data processing pipeline. Experimental findings demonstrate an 11% improvement in fire. Smoke detection accuracy with the horizon clipping algorithm. The successful outcomes using the network were achieved with YOLOv5 reaching an F1 score of 76.75% while processing at a speed of 45 FPS. This provides for detection of fire using these results combined with image enhancement and real time video processing alternatives.

Muhammad et al. [25] applied Laser-Induced Breakdown Spectroscopy (LIBS) and ML techniques for the forensic potential of classifying burnt and unburnt paper samples that are crucial in arson investigations. LIBS was used to analyze various types of paper that were ignited with gas stoves, candles, and lighters. Despite the elemental composition remaining constant among all the ignition types, complexity in spectral data required sophisticated analysis to discriminate between them. In order to deal with this issue, the authors used Principal Component Analysis (PCA) and several supervised learning algorithms that include Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and SVM. Supervised models proceeded to classify 100% of spectrally subtly altered samples achieving 100% classification accuracy, indicating how supervised models have the ability to differentiate spectroscopically between what were often seemingly nearby samples. Finally, this study provides a good example of the success of employing ML techniques for anomaly detection tasks in specific fields such as fire-related document analysis.

Yuan et al. [26] presented a thorough review of chemical analysis of fire debris advancements both in terms of instrumental development and in the emerging data interpretation techniques. The work is on the integration of machine ML to interpret complex chromatographic data using onsite analytical tools, laboratory instrumentation and computational methods. Powerful methods for the detection of ignitable liquids include comprehensive two-dimensional Gas Chromatography coupled to Time-Of-Flight Mass Spectrometry (GC × GC-TOFMS), Direct Analysis in Real Time Mass Spectrometry (DART-MS), and Headspace Gas Chromatography Ion Mobility Spectrometry (HS-GC-IMS). Furthermore, the review emphasizes the role of ML and chemometric approaches in pattern recognition to classify ignitable liquids more accurately especially when pyrolysis products and background volatiles are present. The authors advocate for the validation of such approaches to ensure admissibility in forensic contexts. With the emergence of portable spectroscopic and electrochemical sensor technologies, alongside automated data interpretation via ML, the study suggests a clear trend toward replacing traditional manual methods and even canine detection units in fire debris analysis. This review underscores the shift toward more robust, reliable, and faster forensic anomaly detection methodologies through technological advancement.

The shortcomings of each study will be presented in Table Table 1.

3. Methodology

This section details the methodology employed for anomaly event detection using YOLOv9, specifically focusing on the case study of arson fires. The methodology

Table 1. Summarizing the references and their shortcomings.

Reference	Methodology	Dataset	Results	Shortcomings
[17]	CNN and RNN for real-time detection of suspicious behaviors	UCF-Crime Dataset	Trained six variations of models with adjustments in dataset refinement, regularization, optimizer, loss function, and data augmentation to improve detection accuracy and reduce overfitting.	The approach is generalized to detect aggressive and violent behaviors rather than specifically detect fire or arson incidents.
[18]	CNN and YOLOv9 for real-time weapon detection in videos	Custom Weapon Dataset	97.62% accuracy	Emphasizes CNN and YOLOv9 for weapons, which might not fully address specific fire detection challenges.
[19]	Semi-supervised DL with multimodal data for anomaly detection in sensitive environments like ATMs	Human Action Recognition Dataset in ATM (HARD-ATM).	Competitive with state-of-the-art on benchmark datasets	Limited computational complexity reduction might not scale well to larger datasets required for arson detection.
[20]	CNN-based model for traffic accident detection using the VAID dataset and rolling average prediction	VAID Dataset, 30 surveillance videos	82% accuracy achieved on high resolution videos of traffic.	The performance drops dramatically in poor visibility or distant scenes, and there is relatively small dataset size.
[21]	Insider threat detection using behavioral feature from logs using LSTM autoencoder.	CMU CERT v4.2 Insider Threat Dataset	Accuracy: 90.60%, Precision: 97.00%, F1-score: 94.00%	Focuses on user log data rather than visual data.
[22]	Distributed K-means/RF and DL models (CNN, LSTM) on a two-stage hybrid IDS.	NSL-KDD and CIC-IDS2017	Accuracy: 85.24% on NSL-KDD and 99.91% on CIC-IDS2017	Applied to network traffic rather than physical anomaly.
[23]	ML models (RF, SVM, CNN) for fire residue classification	GC-MS dataset of fire residue data	Accuracies of 0.88–0.92	The models used (random forest, support vector machine, CNN) may not perform optimally for real-time anomaly detection in video streams.
[24]	Drone-based fire detection with YOLO5, horizon line detection, and stabilization techniques	custom forest fire image dataset collected using drones	F1 score of 76.75% with 45 FPS	The horizon line detection algorithm introduced a slight delay.

Table 1. Continued.

Reference	Methodology	Dataset	Results	Shortcomings
[25]	LIBS with ML and PCA for classifying burnt/unburnt paper	A custom experimental dataset based on ignited paper samples.	accuracy: 100%	The method deals with static spectroscopic data that and do not apply in dynamic real time detection systems.
[26]	Review of ML/chemometric techniques for fire debris chemical analysis	Various forensic datasets and instruments (GC × GC-TOFMS, DART-MS, HS-GC-IMS)	identified towards automation and use of ML in forensic analysis.	No direct implementation of model or performance metrics for real-time detection tasks.

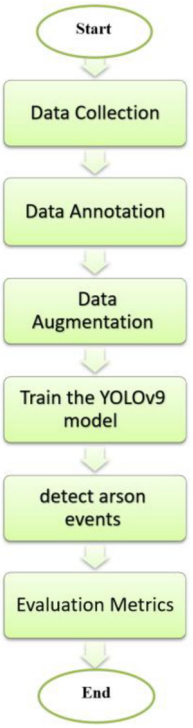

Fig. 1. Flowchart for anomaly event detection using YOLOv9.

encompasses data collection, model training, anomaly detection processes, and evaluation metrics, as illustrated in Fig. 1.

3.1. Dataset description and preprocessing

The preprocessing steps included removing irrelevant data from images and videos and formatting them to a fixed size (640 × 640) to enhance model robustness. In total, 290 frames were manually selected from 53 videos containing arson scenes. The data

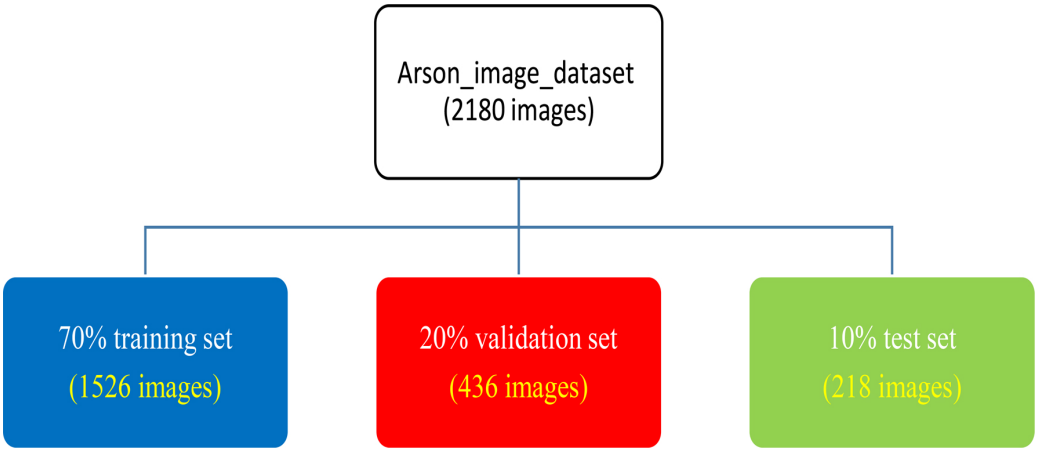

Fig. 2. Splitting of data into three sets: training (blue), validation (red), and test (green).

includes scenes from the anomaly detection dataset [27], which Chen collected from the University of North Carolina at Charlotte. The data was then augmented using a series of transformations to enhance the model's robustness and improve its generalization ability. These transformations were implemented through a Python script designed by the researcher. The augmentation techniques included rotation (up to ± 45°), horizontal flipping (100% probability), random scaling (± 20% of the original size), brightness and contrast adjustments (±20% variation), Gaussian noise addition (variance range: 10–50), Gaussian blurring (blur limit of 3), contrast enhancement (±20% variation), gamma correction (range: 80–120), and hue-saturation adjustments. These transformations ensured variability in the dataset by simulating different lighting conditions, perspectives, and image distortions. As a result, the total number of images increased to 2,182. The data was divided into a training set (70%), a validation set (20%), and a test set (10%) using a Python script designed explicitly for this purpose, as illustrated in Fig. 2.

3.2. Data annotation and set hyperparameters

After preprocessing, the images were then annotated, and bounding boxes were defined according to the labelingImg pro, as illustrated in Fig. 3. The annotations helped accurately mark the regions of interest in each image. This process ensures that the model can effectively learn to identify relevant features in arson scenes. Also, a set of hyperparameters that have been carefully chosen were utilized for appropriately configuring the model training process for optimal training performance. For example, the learning rate of Stochastic Gradient Descent (SGD) is set as 0.01, and the final Learning Rate Factor (LRF) is 0.01. In order to increase the convergence stability, the momentum parameter was set to 0.937. In order to prevent the model from overfitting, a weight decay of 0.0005 was applied to the model. Furthermore, prior to early training stabilization of the process by a warm-up phase was first adopted by implementing three epochs with the initial warm up momentum being 0.8. The ranges of these hyperparameter choices were purposefully chosen to optimize between convergence speed and model generalizability in line with the principles of DL optimization.

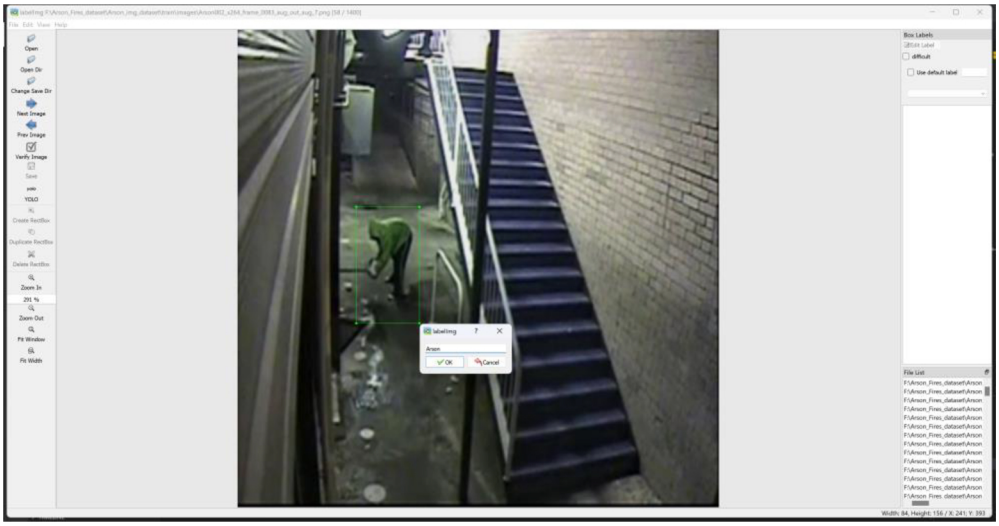

Fig. 3. Data annotation by labelingImg script.

3.3. Model selection and architecture

YOLOv9 was chosen for real-time object detection. Owing to its efficiency and high accuracy, it is appropriate for detecting anomalies in dynamic environments such as arson incidents [28]. YOLOv9 was chosen because it is very efficient and has high accuracy in real time object detection tasks, which are important for detecting dynamic anomalies like arson incidents in surveillance videos. YOLOv9 is unique in that it provides a good balance of speed and accuracy and, therefore, is suited for use when quick responses are required. Moreover, its new architecture includes certain advanced features, such as programmable gradient information (PGI), which enhances the capability of the model to learn and reduces information loss, even in complex environments. Because of these attributes, YOLOv9 is a strong candidate for arson detection, as it can process data quickly and accurately (e.g., far more quickly and, thus, far less adaptively than Faster R-CNN and SSD). PGI is used to control the propagation of gradient information for different semantic levels in the YOLOv9 architecture. The primary branch, auxiliary reversible branch, and multi-level auxiliary information make up PGI. It only uses the main branch, which manages forward and backpropagation during inference. An information bottleneck could develop as the network gets deeper, resulting in loss functions that cannot generate helpful gradients. Under such circumstances, the auxiliary reversible branch uses reversible functions to reduce information loss in the main branch and maintain information integrity. Furthermore, multi-level auxiliary information improves the model's learning ability by introducing supplementary information at various levels and addresses the problem of error accumulation from the deep supervision mechanism. This is the theoretical foundation of the YOLOv9 model's superior performance in arson detection tasks, as shown in Fig. 4 [29].

The YOLOv9 model used for arson detection operates with approximately 57.3 million parameters. This highly efficient architecture requires this parameter count to support its high accuracy and real-time processing speed, which makes it practical for dynamic tasks like detecting anomalies in arson incidents. Combining these parameters with the PGI framework further enhances its capability to handle complex environments with minimal

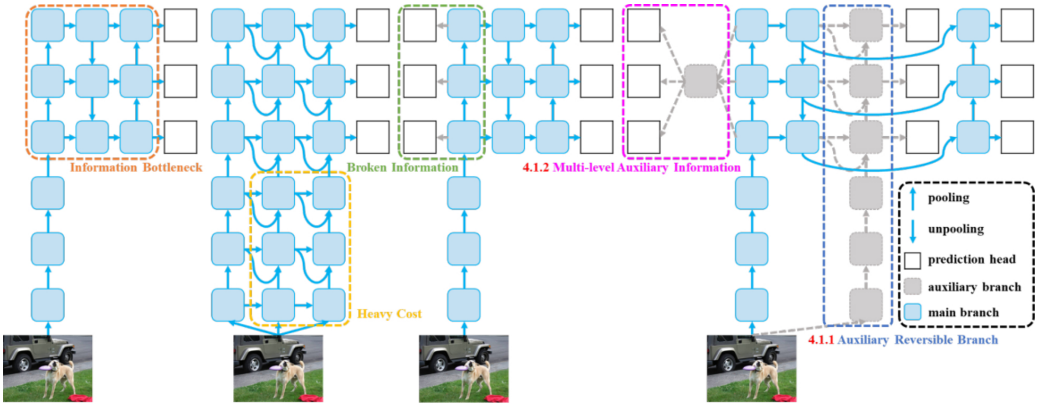

Fig. 4. YOLOv9 architecture [28].

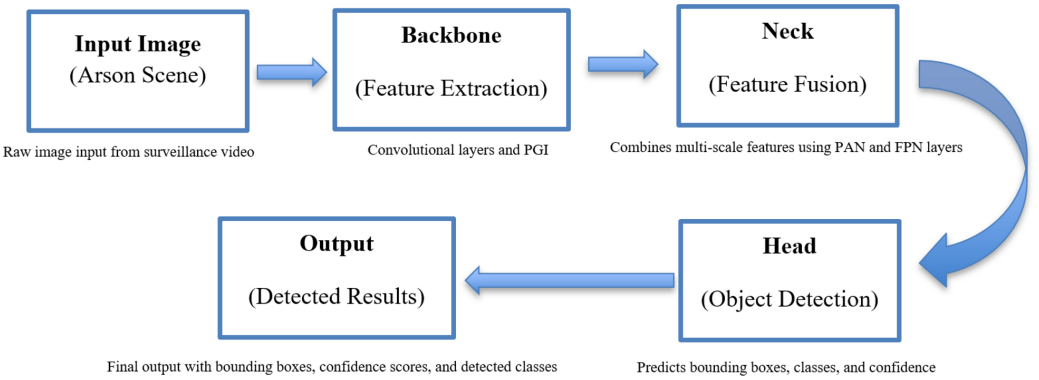

Fig. 5. Diagram of YOLOv9 for arson detection.

information loss across deep layers [28]. Fig. 5 provides a visual representation of the YOLOv9 architecture used in this study and highlights the data flow through its main components: the input image, the backbone to extract the feature, the neck to fuse the features, and the head to detect objects. This design integrates PGI to bolster the learning characteristics and guarantee that little data is lost in deep layers of the architecture, making it very effective in detecting arson.

3.4. Model evaluation and performance metrics

To evaluate the performance of the YOLOv9 model in Arson_img_dataset, the following metrics were used:

– Precision is the proportion of correctly predicted positive observations to all predicted positives [29].

$$Precision = \frac{TP}{(TP + FP)} \tag{1}$$

where TP is the True Positives and FP the False Positives.

– Recall is the proportion of all observations in the actual class that were accurately predicted to be positive [29].

$$Recall = \frac{TP}{(TP + FN)} \tag{2}$$

where TP is the True Positives and FN is the False Negatives
– F1-score is the weighted average of recall and precision [30].

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{3}$$

– mean Average Precision (mAP) is the average precision score for each class. Average Precision (AP) is calculated for each class, and mAP is the mean of these values [31–33].

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{4}$$

4. Results and discussion

The experimental evaluation demonstrates the effectiveness of YOLOv9-small (YOLOv9-s) in arson detection across multiple operational scenarios. The YOLOv9-s model achieved robust performance despite computational challenges during initial training. Section 4.1 describes the training process in detail, Section 4.2 outlines the quantitative performance analysis, and Section 4.3 provides a comparison with baseline models.

4.1. Training setup and select epoch

The data was trained on a personal computer with specifications (CPU cori7, ram 16 Gb), and the training failed because it took a very long time (up to 25 days). Therefore, Google Colab was used with certain specifications (A100 GPU Ram 40 Gb, System Ram 80 Gb, Disk 200 Gb) when the training was done in two cases (55 and 100 epochs), and the evaluation of the two cases indicated that 100-epoch showed overfitting. Therefore, the 55-epoch case was used because it yielded better results. The training period lasted 30 minutes using Google Colab, whose specifications are mentioned above. The model summarizes the training case, as shown in Fig. 6.

4.2. Evaluation metrics and results

The results of the tests were analyzed to shed light on how well the model performed in identifying incidents of arson as in Table 2. The mAP for the model was 0.552 at a threshold of 0.5, which shows that it balances accuracy and completeness. Arson can be difficult to detect but can have very serious consequences if it is not. A high mAP value indicates the model's strong ability to differentiate between instances of arson and non-arson. According to the precision confidence curve, the precision reaches 1.00 when the confidence level reaches 0.933, highlighting the model's accuracy when it is confident in its predictions. This high level of precision suggests that the model may occasionally mislabel something as arson when it is very particular, making it a reliable tool for identifying arson incidents during moments that demand very high accuracy. Likewise, the recall confidence curve

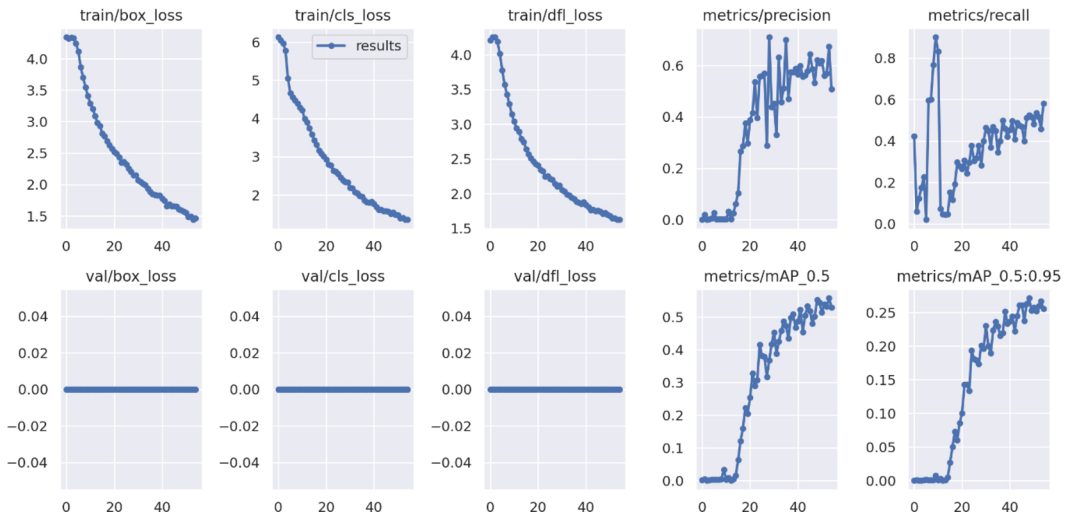

Fig. 6. Training results.

Table 2. Performance results of YOLOv9s with arson dataset.

Model	Precision	Recall	F1-score	mAP@0.50	mAP@0.50:0.95	Training time (Hr)
Yolov9-s	0.445	0.507	0.560	0.456	0.219	3.191

shows that the model can identify some of the actual arson cases at these confidence levels (recall rate = 0.84). This shows that the model is robust and sensitive to capture instances, which are necessary for identifying and responding to incidents. The F1 score enables evaluation by considering both precision and recall aspects. The peak F1 score of 0.58 at a confidence level of 0.39 indicates the model's effectiveness and high performance at confidence levels. For practical use, it is important to find the best mix between detecting and reducing unnecessary alerts. The results demonstrate the performance of the model in the presence of high accuracy and high recall. Nonetheless, there is still room for enhancement concerning F1 score and sustaining performance across varying confidence levels. In future, efforts could be made to tune the model and explore ways to improve data quality for precision and dependability.

The validation batch's evaluation results, which highlight the arson incident detection and confidence levels, are shown in Fig. 7. The model demonstrates strong detection capabilities with varying confidence levels, reflecting its training efficiency and accuracy.

Fig. 8A illustrates the precision-recall curve for the arson category, illustrating how the model trades off precision and recall at different confidence thresholds. With a value of 0.5, the model achieves a mAP of approximately 0.552, thus indicating that it detects arson while performing reasonably well in terms of false positives. The model has perfect precision (1.00 $\pm$ 0.02 at a confidence of 0.933) at high confidence levels (above 0.7), although this precision comes at the expense of recall (0.18 $\pm$ 0.04). Thus, although the model is extremely accurate in its most well-confirmed predictions, it can detect only a small fraction of arson incidents. Automated alert systems are one example of a case in which such high precision is advantageous, but the low recall means that further detection mechanisms are required for comprehensive coverage. In contrast, with a moderate confidence threshold (0.3–0.5), the model shows a more balanced tradeoff between precision and recall, with an F1-score of 0.58 $\pm$ 0.02 at a confidence level of 0.39. Precision stabilizes

Fig. 7. Validation batch evaluation results.

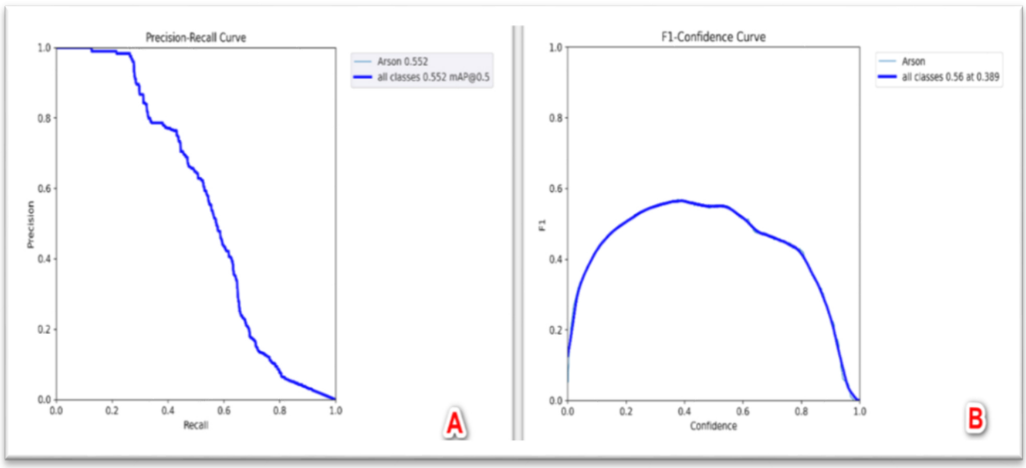

Fig. 8. (A) Precision–recall curve and (B) F1 score vs. confidence curve.

to 0.47 $\pm$ 0.03, and recall stabilizes to $\approx$ 0.61 $\pm$ 0.05; this equilibrium is operationally effective in general surveillance. At lower confidence thresholds (below 0.2), the model prioritizes recall (0.84 $\pm$ 0.03) over precision (0.22 $\pm$ 0.02) since it tends to report most true arson cases with a greater number of false positives. The model is not appropriate for fully autonomous systems, but it works well as an initial screening step in a multi-stage detection where subsequent verification stages can remove erroneous detections. Fig. 8B presents the F1 score as a function of the confidence threshold, providing insight into the

Fig. 9. (A) The heatmap of the original image and (B) The heatmap generated by the YOLOv9 model.

tradeoff between precision and recall across different confidence levels. This is a diagnostic tool for discovering the threshold that gives the model the best tradeoff between these two metrics to increase its overall detection reliability. The F1 score for the arson category reaches its peak at around 0.39 and declines gradually as confidence rises. This indicates that the system balances precision and recall at confidence levels, achieving a maximum F1 score of approximately 0.58 and a confidence level near 0.39. The system's overall performance is also reflected in an F1 score of 0.56 at a confidence level of 0.389 across all categories.

Visual explanation techniques such as heatmaps were employed to assess the interpretability and effectiveness of the model. These tools make the model's predictions more transparent and trustworthy by showing which parts of the image the model pays attention to when making predictions. A heatmap transforms complex data into a vibrant, color-coded matrix [34]. The original image in Fig. 9A has no heatmap, while Fig. 9B illustrates the heatmap generated by the YOLOv9 model, in which more attention is concentrated on specific regions of the image, particularly the area around the fire. In the context of arson detection, the YOLOv9 model's decision-making process can be analyzed by interpreting its attention through these heatmaps. As fire detection requires a model to distinguish between accidental fires and deliberate arson fires, the heatmaps show which features—such as human behavior, fire spread patterns, or unusual objects near the fire—are deemed most relevant. The red and yellow regions in the heatmap correlate with the features that the model associates with arson, confirming that it focuses on the image's critical elements that indicate intentional fire-setting behavior. Visualization tools like heatmaps allow researchers and practitioners to understand the underlying reasons for a model's decisions, increasing confidence in its ability to detect arson and prevent false positives. However, these interpretability techniques can assess the model's performance in different scenarios. Depending on the situation, the model should concentrate more on human interactions and behaviors and less on suspicious activities in the case of a purposeful fire and vice versa in the case of an accidental fire. This flexibility in the model's interpretability is crucial for ensuring accurate and reliable arson detection in real-world environments.

Table 3. Comparison of optimized model (Model 6) and YOLOv9-s.

Aspect	Optimized model (model 6) [17]	Proposed model
Model Name	Model 6	YOLOv9-s
Number of Categories	13 (Abuse, Burglar, Explosion, Shooting, Fighting, Shoplifting, Road Accidents, Arson, Robbery, Stealing, Assault, Vandalism, Normal)	1 (Arson)
Optimizer	Stochastic Gradient	N/A
Loss Function	Categorical cross-entropy	N/A
Regularization	l2 (0.01)	N/A
Activation Functions	Rectified Linear Unit (ReLU), Sigmoid, SoftMax	N/A
Augmentation	Horizontal Flip	N/A
Precision	N/A	0.445
Recall	N/A	0.507
F1-score	N/A	0.560
mAP@0.50	N/A	0.456
mAP@0.50:0.95	N/A	0.219

4.3. Performance comparison with other models

A comparative analysis of the optimized model (Model 6) and the proposed YOLOv9-s model is presented in Table 3. With more characteristics, the optimized model (Model 6) recognizes 13 categories of anomalies, including abuse, burglary, explosion, and arson, while YOLOv9-s detects only arson cases. While Model 6 uses the stochastic gradient optimizer and categorical cross-entropy loss with a 12-regularization type, the efficiency of this YOLOv9-s is optimized for a single type of anomaly. Also, Model 6 uses various data augmentation techniques like horizontal flipping to strengthen the image set. Nevertheless, in Model 6, the degree of model complexity has been improved, but there is no presentation of the current analysis of precision, recall, and F1-score indicators. On the other hand, the proposed YOLOv9-s model provided a precision of 0.445, recall of 0.507, and F-score of 0.560, with good results, particularly in arson detection. For the mAP@0.50 and mAP@0.50:0 under the threshold type H, the accuracy values of 0.456 and 0.219 observed for the model on 95 deny further stress the model's availability under different thresholds. This comparison shows that the nature of the anomaly is crucial for structuring the model or choosing the optimal optimizers. Further research could explore how to combine a generalization type of model, such as Model 6, with fine-tuned models, such as the YOLOv9-s, to provide fair generalization and accuracy for different real-life conditions.

5. Conclusions and future work

This study aimed to create and assess a system for detecting instances of arson using the YOLOv9 learning model in surveillance videos. The research gathered and refined data labeling, enhanced data, trained the model, evaluated its performance, and compared the results while considering quality, diversity, and environmental factors. The model was able to achieve a precision (mAP) of 0.552 with a threshold of 0.5, meaning that precision and recall are balanced in the sense of detecting arson incidents. The model was found to be robust and sensitive to arson cases with precision of 0.933 and recall rate of 0.84 at a confidence level of 0.0. The F1 score at the peak of around 0.58 at a confidence level of 0.39 indicates that the model performs the best at this confidence level. The results indicate that the model can determine with certainty whether a case is arson and can also find arson cases at low confidence levels. It is important to balance accuracy and coverage since mistakes can be large. Nevertheless, some work still needs to be done to achieve F1 scores and stable performance across a wide range of confidence levels. Dealing with

data quality and various environmental conditions enhances the model's robustness and dependability.

The current study is suggested to be expanded by the following areas for exploration:

- Generative adversarial networks (GANs) can also help increase dataset diversity by generating arson related images and thereby reduce the need for collecting too much hands on data. GANs help the model generalize better to real-world scenarios by simulating variations in fire intensity, smoke patterns, and lighting conditions. However, the careful balancing of real and synthetic data is critical to prevent overfitting and building a robust arson detection model that can handle diverse conditions.
- Hyperparameter tuning of the YOLOv9 model for improving model performance using optimization algorithms like particle swarm optimization (PSO), genetic algorithm (GA), and grid search (GS), random search (RS) and find better performance metrics rate like mAP & F1 scores. These algorithms select optimum hyperparameters to fine-tune the model's architecture and the training process, which helps enhance detection accuracy as well as generalization to different arson detection scenarios.
- Improving model performance through the hyperparameter tuning of the YOLOv9 model using optimization algorithms such as PSO, GA, GS, and RS. With these algorithms, the model will find the best hyperparameters. This approach is expected to improve the model's generalization to different arson detection scenarios and improve its detection accuracy.
- Increasing the amount of detection precision and learning more about incidents across the board through integration with additional systems by incorporating sensors such as cameras or audio detectors. Additionally, developing robust alert systems and user interfaces would ensure that authorities are promptly and accurately alerted to any detected anomalies.

Acknowledgment

I would like to thank Mr. Mahmood A. Jumaah for his invaluable technical support that made it possible for me to complete this study.

Conflict of interest

The authors declare no conflict of interest.

Authors contributions

The draft was prepared by the first author, Ali Abbas Abbod, the review, editing was performed by the second author, Matheel E. Abdulmunim, where the final revision of the whole manuscript, including filling in technical details was undertaken by the third author, Ismail A. Mageed.

Data availability

The anomaly datasets are available on https://www.kaggle.com/datasets/minhajuddinmeraj/anomalydetectiondatasetucf.

References

1. M. Khanam, "A fully online approach for anomaly detection and change-point detection in streaming data using LSTM," Ph.D. dissertation, Department of Computer Science, University of Manchester, Manchester, England, 2023.
2. R. Foorthuis, "On the nature and types of anomalies: a review of deviations in data," *International journal of data science and analytics*, vol. 12, no. 4, pp. 297–331, Aug. 2021, doi: 10.1007/s41060-021-00265-1.
3. X. Serrano-Guerrero, G. Escrivá-Escrivá, S. Luna-Romero, and J.-M. Clairand, "A time-series treatment method to obtain electrical consumption patterns for anomalies detection improvement in electrical consumption profiles," *Energies*, vol. 13, no. 5, Art. no. 1046, Feb. 2020, doi: 10.3390/en13051046.
4. Y. Liu *et al.*, "Generalized video anomaly event detection: Systematic taxonomy and comparison of deep models," *ACM Computing Surveys*, vol. 56, no. 7, Art. no. 189, April 2024, doi: 10.1145/3645101.
5. J. Liu *et al.* "Networking systems for video anomaly detection: A tutorial and survey," 2025, *arXiv:*2405.10347.
6. W. Ullah *et al.*, "Artificial intelligence of things-assisted two-stream neural network for anomaly detection in surveillance big video data," *Future Generation Computer Systems*, vol. 129, pp. 286–297, April 2022, doi: 10.1016/j.future.2021.10.033.
7. M. Firlej and A. Taeihagh, "Regulating human control over autonomous systems," *Regulation & Governance*, vol. 15, no. 4, pp. 1071–1091, July 2021, doi: 10.1111/rego.12344.
8. N. H. Ali, M. E. Abdulmunem, and A. E. Ali, "Learning evolution: A survey," *Iraqi Journal of Science*, vol. 62, no. 12, pp. 4978–4987, Dec. 2021, doi: 10.24996/ijs.2021.62.12.34.
9. N. Wang, S. Zhao, S. Cui, and W. Fan, "A hybrid ensemble learning method for the identification of gang-related arson cases," *Knowledge-Based Systems*, vol. 218, Art. no. 106875, April 2021, doi: 10.1016/j.knosys.2021.106875.
10. T. Diwan, G. Anirudh, and J. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, Aug. 2022, doi: 10.1007/s11042-022-13644-y.
11. I. Syed, A. S. K. Hasane, K. A. Sai, B.V., G. D., and R. V, "YOLO (YOU ONLY LOOK ONCE) making object detection work in medical imaging on convolution detection system," *International Journal of Pharmaceutical Research*, vol. 12, no. 2, pp. 312–326, 2020, doi: 10.31838/ijpr/2020.12.02.0003.
12. A. C. Bukola, P. A. Owolawi, C. Du, and E. V. Wyk, "A systematic review and comparative analysis approach to boom gate access using plate number recognition," *Computers*, vol. 13, no. 11, Art. no. 286, Nov. 2024, doi: 10.3390/computers13110286.
13. C. Taylor, M. Widjaja, and O. D. Dantsker, "Remotely-processed vision-based control of autonomous lighter-than-air UAVs with real-time constraints," In *Proc. of the AIAA Science and Technology Forum and Exposition Int. Conf*, Orlando, FL, USA, 6–10 Jan. 2025, doi: 10.2514/6.2025-1344.
14. A. A. Alsabei, T. M. Alsubait, and H. H. Alhakami, "Enhancing crowd safety at hajj: Real-time detection of abnormal behavior using YOLOv9," *IEEE Access*, vol. 13, pp. 37748–37761, 2025, doi: 10.1109/ACCESS.2025.3545256.
15. T. Oldag and T. Enright, "Fire investigation overview," in *The Path of Flames*, 1st ed. UK: CRC Press, ch. 8, pp. 116–130, 2024.
16. M. W. K. S. Durage, "UAV-based wildfire analysis," Master thesis, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland, 2024.
17. V. Singh, S. Singh, and P. Gupta, "Real-time anomaly recognition through CCTV using neural networks," *Procedia Computer Science*, vol. 173, pp. 254–263, 2020, doi: https://doi.org/10.1016/j.procs.2020.06.030.
18. A. M. Hundalekar, V. Shirsath, V. R. Naidu, and S. B. Pingale, "Intelligent monitoring for anomaly recognition using CNN and YOLOv9," *International Journal of Scientific Research & Engineering Trends*, vol. 10, no. 2, pp. 744–749, 2024.
19. P. Khaire and P. Kumar, "A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments," *Forensic Science International: Digital Investigation*, vol. 40, Art. no. 301346, Mar. 2022, doi: 10.1016/j.fsidi.2022.301346.
20. S. W. Khan *et al.*, "Anomaly detection in traffic surveillance videos using deep learning," *Sensors*, vol. 22, no. 17, Art. no. 6563, Aug. 2022, doi: 10.3390/s22176563.
21. R. Nasir, M. Afzal, R. Latif, and W. Iqbal "Behavioral based insider threat detection using deep learning," *IEEE Access*, vol. 9, pp. 143266–143274, 2021, doi: 10.1109/ACCESS.2021.3118297.
22. C. Liu, Z. Gu, and J. Wang "A hybrid intrusion detection system based on scalable k-means+ random forest and deep learning," *IEEE Access*, vol. 9, pp. 75729–75740, 2021, doi: 10.1109/ACCESS.2021.3082147.
23. C. Park, J.-B. Lee, W. Park, and D.-K. Lee, "Fire accelerant classification from GC–MS data of suspected arson cases using machine–learning models," *Forensic Science International*, vol. 346, Art. no. 111646, May 2023, doi: 10.1016/j.forsciint.2023.111646.

24. N. Abramov *et al.*, “Intelligent methods for forest fire detection using unmanned aerial vehicles,” *Fire*, vol. 7, no. 3, Art. no. 89, Mar. 2024, doi: 10.3390/fire7030089.
25. N. Muhammad, M. Faheem, M. Manzoor, A. Gulzar, M. Bilal, and Y. Jamil “Machine learning assisted libs classification of burnt and unburnt paper samples: A forensic perspective.” *Talanta*, vol. 293, Art. no. 128073, Oct. 2025, doi: 10.1016/j.talanta.2025.128073.
26. Y. Low, E. Tyrrell, E. Gillespie, and C. Quigley, “Recent advancements and moving trends in chemical analysis of fire debris,” *Forensic Science International*, vol. 345, Art. no. 111623, April 2023, doi: 10.1016/j.forsciint.2023.111623.
27. A. Singh, A. Bajaj, A. Singh, S. D. Saha, and A. Sharma, “Exploring anomaly detection techniques for crime detection,” In *Proc. of the Int. Conf. on Recent Trends in Communication & Intelligent Systems (ICRTCIS 2024)*. Singapore, pp. 183–201, doi: 10.1007/978-981-97-7632-0_13.
28. C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “YOLOv9: Learning what you want to learn using programmable gradient information,” 2024, *arXiv: 2402.13616*.
29. D. R. Chirra, “Deep learning techniques for anomaly detection in IoT devices: Enhancing security and privacy,” *Revista De Inteligencia Artificial En Medicina*, vol. 14, no. 1, pp. 529–552, 2023.
30. R. Raturi, A. Kumar, N. Vyas, and V. Dutt, “A novel approach for anomaly detection in time-series data using generative adversarial networks,” In *Proc. 2023 Int. Conf.* on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, pp. 1352–1357, doi: 10.1109/ICSCSS57650.2023.10169365.
31. G. Kaur and S. Saini, “Comparative analysis of RMSE and MAP metrics for evaluating CNN and LSTM models,” in *Proc. Int. Conf. Recent Advancements in Communication, Computing, and Artificial Intelligence (RACCAI-2023)*, Mohali, India, Art. no. 040003, doi: 10.1063/5.0221565.
32. J. C. Obi, “A comparative study of several classification metrics and their performances on data,” *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, 2023, doi: 10.30574/wjaets.2023.8.1.0054.
33. Y. Ge, Q. Liu, G. Wu, S. Wang, and F. Ye, “Score-ranking smooth average precision loss for remote sensing image retrieval,” *Journal of Spatial Science*, vol. 69, no. 3, pp. 801–820, Jan. 2024, doi: 10.1080/14498596.2024.2302167.
34. Dinov, Ivo D. “Basic visualization and exploratory data analytics,” in *Data Science and Predictive Analytics*, 2nd ed. USA: Springer, Cham, ch. 2, pp. 61–147, 2023, doi: 10.1007/978-3-031-17483-4_2.